DOCUMENT RESUME

ED 040 124 SP 003 856

AUTHOR Herbert, John
TITLE A Research Base for the Accreditation of Teacher Preparation Programs.
PUB DATE 70
NOTE 34p.; Paper presented at annual meeting, AERA, Minneapolis, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.80
DESCRIPTORS *Accreditation (Institutions), *Evaluation Criteria, *Followup Studies, Program Evaluation, *Research Needs, Standards, Teacher Behavior, *Teacher Education, Teacher Persistence

ABSTRACT

In establishing a research base for the accreditation of teacher preparation programs, the present standards, as established in the "Recommended Standards for Teacher Education" (ED 037 423) or the American Association of Colleges for Teacher Education (AACTE), should be developed into a set of multiple standards to fit diverse programs and several levels of quality. Additional criteria for measuring the effects of preparation programs should also be formulated, based on career line information, especially retention in teaching, client satisfaction, and above all, the teaching behavior of students during the program and after graduation. Typical problems encountered in reviewing research on the relationship of teacher behavior to preparation programs are the lack of replication of studies, the lack of information given on specific research procedures, and the lack of a common theoretical framework. One step toward overcoming these problems would be the establishment of an evaluation team for screening research relevant to teacher education. Organizations like NEA and AERA could cooperate to develop standardized research designs to be made available to teacher preparation programs. (RT)

ED040124

DRAFT
Not For Publication

A Research Base For The Accreditation Of Teacher Preparation Programs*

John Herbert, The Ontario Institute For Studies In Education and
The University of Toronto

The other papers of this symposium ask what light research has thrown and can throw on the present criteria, procedures, and standards of accreditation of basic teacher preparation programs adopted by the American Association of Colleges For Teacher Education on the recommendation of its Evaluative Criteria Study Committee. It is the purpose of this paper to consider how a research base might be established for the development of alternative or supplementary accreditation standards. Such research would deal with questions of curriculum evaluation and design, and with the evidence we have and need in guiding institutions in strengthening their teacher preparation programs.

The Recommended Standards

The direction here recommended is in keeping with the current policy of the AACTE as expressed in the new Recommended Standards For Teacher Education (AACTE, 1969). While previous drafts were aimed to "help to protect children and youth from ill-prepared school personnel" (AACTE, 1968, p. 1), the newest document

*A report prepared for the Special Interest Group on the Teacher Preparation Curriculum and Division B, AERA, Minneapolis, 1970

SP003856

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

clearly states that the goal is to set up procedures which will assure the public that accredited programs "meet national standards of quality," that "children and youth are served by well-prepared personnel," and that the teaching profession is advanced "through the improvement of preparation programs." (AACTE, 1969, p. 1) While the earlier goal merely made it necessary to identify the bad eggs, the new aims require a much clearer knowledge than we now have of the possible meaning of the word "standards" in the current document, the relationship between the nature of programs and the teaching ability of their graduates and the values which should inform efforts toward improving programs. The changes here proposed are also in accordance with some of the most advanced proposals for changes in teacher preparation. (Stiles, 1968,) Before turning to the question of alternative criteria and research evidence, I should like to clarify my position by examining some underlying issues raised by the Recommended Standards.

Criteria, Standards, and Values

In discussing these issues it will be helpful to make a distinction between two words which are often used synonymously: "criteria" and "standards." I will use the word "criterion" to refer to a characteristic which is to be examined by an accrediting team. I will reserve the word "standard" for a qualitative or quantitative measure of the degree or extent to which a program

possesses that characteristic. For example, the following statement in section G4.1 of the Recommended Standards is by my definition a criterion: "Standard: the library is adequate to support the instruction, research and services pertinent to each teacher preparation program." (AACTE, 1969, p. 11) The subsections, which consist of questions pertaining to such matters as diversity of holdings, library use, and annual expenditure, make this criterion more specific, so that the staff of the program knows what characteristics of the library the accrediting team will consider. But only if a minimum standard is explictly stated (e.g., the library shall contain at least 200 dollars worth of books per student), can the program see how far it must improve its library to attain accreditation. And only if a continuum or set of continua is presented, indicating various standards below and above the minimum for each specific criterion, can a program compare its resources with those of other programs or aim at a given degree of improvement.

With very rare exceptions, the present document indicates the general areas of a program to be assessed - that is, it establishes criteria - but it does not state standards. One exception is the requirement that at least one-third of any program must be in liberal studies. Even here, however, it is unclear whether there is also a top limit which in turn constitutes a minimum for other components. It may be argued that the liberal studies

component is limited by the requirements of the professional component, but this is described as intended to "provide a set of categories through which an institution can describe and review the professional studies component of the various teacher education curricula it offers." So the requirement here is merely that each program will contain something recognizable as belonging to each named category.

What this amounts to is a set of criteria analogous to those established for the evaluation of libraries. Such criteria might be thought of as minimum standards: a program must show some evidence of attention to each of the criteria in all the categories in the document. If this is the case the standards are so low that they are unlikely to serve as incentives for improvement and will at best duplicate state and regional accreditation. More likely, however, there are hidden standards behind the criteria, or each accrediting team must in practice establish its own standards, adjusting them perhaps to the professional goals and values of particular programs it is responsible for examining.

The AACTE has thus tacitly recognized two serious issues: the problem of conflicting values within or between accreditation groups and between accreditation groups and the program to be accredited; and the difficulty of obtaining evidence adequate to establish or support standards. Accreditation, even when it is

intended merely to ensure minimum resources or continued progress towards any goals a program may set for itself, nevertheless assumes that some values exist. These values may be economic: maximum allocation or optimum use of facilities or of human and curricular resources. They may be pedagogical: preparing the type of teacher held to be most effective in certain kinds of schools. Or they may simply be developmental: assuring that programs continuously evaluate and modify their own practices. In any case the determination of such values precedes accreditation procedures, even the establishment of criteria. To list criteria but to leave standards inexplicit does not in itself resolve the problem of conflicting or unsubstantiated values.

Cut-off Points

Research based on the current Recommended Standards might help to clarify accreditation procedures and their underlying values by finding out what standards accreditation teams apply in practice, where the cut-off points actually are, how they vary for different types of programs, and how much agreement or disagreement arises within and between accrediting teams and between accrediting teams and program staffs. We could also seek to determine whether teacher preparation programs or their staffs show any detectable change when these minimum standards are applied. (Brase, 1964)

It seems probable, however, that the results of such research would not go very far to implement the stated goals of the present

accreditation procedures. If the minimum standards applied by accreditation teams are low, then only inferior programs will be affected. If the cut-off points are high, then marginal programs may seek to meet them, but programs which are clearly below or above the standard are unlikely to be much affected. A procedure other than the establishing of a cut-off point seems to be required to establish standards beyond the minimum level which simply ensures that an institution has the resources and facilities necessary to operate a teacher preparation program at all.

Single versus Multiple Standards

Obviously, if standards are to exceed such a minimum level, there will be a conflict between the desire to apply a single set of standards to all institutions and so "ensure national standards of quality" (AACTE, 1969), and the desire to leave institutions free to design their own programs. As has been pointed out, "the single standard necessitates the framing of component criteria in very broad terms in order that they be operable," and as a result they are "often only statements of good intention...." (Evers, 1967, p. 61) Again, research may help. It would be possible to describe existing programs of various types in terms of the component criteria provided in the Recommended Standards and to formulate sets of explicit standards for each component of all the main types of program. The components of every program

selected for study could then be rated unsatisfactory, minimal, good, or excellent by standards applicable to programs of that type. The resulting descriptions and ratings, presented as anonymous case studies, could be matched with those of almost any program to be accredited and a corresponding profile identified. As new programs are devised, appropriate descriptions and standards could be added. Instead of a cut-off point, this procedure would employ a series of descriptive-evaluative statements of a number of key components of teacher education programs. It would, moreover, avoid the need to value one educational philosophy over another.

Self-Evaluation

Multiple standards developed in this way would provide a means for detailed assessment and comparison of the facilities, organization, and curricular and human resources of preparation programs, since these are the components dealt with in four of the five sections of the Recommended Standards. The fifth section, "Evaluation, Program Review, and Planning," also provides a basis for describing resources of a somewhat different kind: the institution must conduct "a well-defined plan for evaluating the teachers it prepares" and must use the evaluation results "in the study, development, and improvement of its teacher education programs." The criteria are the existence and employment of evaluation procedures and plans for modification of programs. The institution and not the accrediting agency is to evaluate the

effectiveness of its programs and of its efforts at improvement.

These criteria are valuable in themselves insofar as they indicate a necessary component of preparation programs. Self-evaluation is a continuing process while accreditation is an occasional procedure. In stressing self-evaluation, moreover, the AACTE is in effect giving credit to programs which have conducted or supported research on teaching behavior. As I shall try to show later, there are several ways in which research can help to implement this recommendation. However, it is difficult to see how, finally, an accreditation agency can develop standards for assessing the adequacy of self-evaluation procedures without at some stage making some judgment of the impact of the total program on the teaching behavior of its graduates and on what goes on in the schools. The emphasis on resources is a tradition which goes back at least to the Flowers Report of 1948; but, it has been pointed out, although it was good in its day, "excellence demands more vigorous research in the future particularly on the results the programs achieve." (Mauker, 1962, p.7)

What Should Be Researched?

The question then arises in what way the results of teacher preparation can be assessed - and what kinds of results should be selected for examination. It has often been argued that the validity of the evaluation of a teacher preparation program

increases if the evidence is collected as close as possible to the "final product" or the "third level" - the changes in the pupil. (Woodring, 1957, p. 62) Though indisputable in theory, this argument does not work in practice. While we should do more and better research on which teacher behaviors result in changes in pupil behavior, it is not expedient to evaluate teacher preparation programs by such changes in the schools where the teachers find employment.

Pupil changes occur to a great number of different individuals, each of unknown personality, unpredictable cultural conditioning, and idiosyncratic response. The reaction to any teacher cannot necessarily be attributed to the teacher and much less to the teachers' preparation. Moreover, pupil changes, except responses to tests, are extremely difficult to record accurately. In any case, such changes occur in environments where the teachers of teachers control only one of the variables - the training the teacher receives - and there is evidence that any effect of the training can be driven underground at least temporarily by the anxieties inherent in beginning teaching.

Combination of variables - the school and home environment of the pupils, the decisions of the teacher's peers and administrators, and those of the teacher himself - may result in placing him in a position where regardless of the training received or the criteria

used, he either cannot fail or cannot succeed. It would thus be no more reasonable to evaluate a teacher preparation program by the way pupils learn in the classroom of graduates then to evaluate a program of medical training by the health of the population its graduates serve. Therefore, though it is theoretically attractive to relate pupil behavior to accreditation, this seems unlikely to be feasible in the foreseeable future. As Ryans found: "With all the attractiveness of judgement of teacher behavior from its products [e.g. pupil changes]... the disadvantages of such approaches seem to outweigh their advantages." (Ryans, 1960, p. 71)

When we concentrate instead on teaching behavior the chances of obtaining meaningful information become much greater. The available research is growing rapidly and is already having an impact on teacher preparation. (Bruce, 1969, p. 415) We can use the results of direct observation of teaching and also data about indirect variables which may be related to the teachers' preparation. Both are potentially very fruitful lines of evidence for the accreditation of programs and for the improvement of teacher preparation.

New Complementary Criteria

The criteria which I should like to propose differ from those included in the Recommended Standards in that they are based on the description, not of programs, but of the behavior

of teachers prepared by the programs to be accredited. To establish standards based on these criteria it would not be necessary to establish an invariant relation of particular components of the preparation programs to the behavior of its graduates, since similar results may be produced by disparate causes. Moreover, a program which has achieved results held to be acceptable or desirable should in the absence of strong counter-evidence, be presumed to employ appropriate means. Programs seeking to improve their effectiveness could act on the information gathered during accreditation by attempting to determine specific factors within or beyond their institutions which might affect the teaching behavior of their graduates in desirable and undesirable ways.

Evidence on the behavior of teachers may be gathered by examining records, by obtaining testimony from students, graduates, or supervisors in oral or written form, and by observation of teaching. There are indications that at least some of these types of evidence discriminate among teacher preparation programs (Start, 1967, Report 2, Bledsoe, 1967), but that each is subject to some limitations which would need to be taken into account in formulating criteria and standards. Three of the most promising types of evidence seem to be career line data, client satisfaction, and direct evidence about teaching. Although only a few studies can be cited here, much more research has

been done in each area and there are obvious practical reasons why we should consider each area carefully.

Career line data

Information about career lines includes such matters as wastage from teaching, ratings and recommendations of supervisors, types of teaching and administrative positions held, participation in research and program development, further training and education undertaken, and so on. Some career line information, for example wastage from teaching, would clearly be highly relevant and important for accreditation. The present criteria do not call for any information on the number of years that graduates spend in teaching, only on whether or not they enter the teaching profession (AACTE 1969). It might, however, be considered that an average of at least three years teaching is necessary to justify the expenditure of resources in teacher training. After all, a teacher who remains in teaching for four years costs half as much to educate as two teachers who stay in the profession for two years each. To make wastage a criterion would implement the goals of accreditation since reduction of wastage would cut down the number of inexperienced teachers in the schools and the number of students in preparation programs, thus making for more stable teaching staffs and releasing resources for the improvement of programs.

Information on wastage is already being collected in some states (Charters, 1969, Orlick, 1965) and also by some teacher preparation programs in their follow-up studies reported in the educational journals. Studies of teacher mobility often omit information on training programs in preference to information on age and sex, (NEA, 1969) which of course are clues to the incidence of marriage and pregnancy, two major reasons why teachers drop out, but which are probably of less significance to the profession. Researchers may be able to describe groups or types of graduates and the conditions under which they have high or low survival rates. They could investigate variables which seem likely to be related to wastage, including the appropriateness of the new teachers' skills for the initial teaching position, and attempt to locate particular program components or variables which might be altered to reduce wastage. Such research would be of considerable practical and theoretical interest.

Other kinds of career line data, though from a commonsense position they <u>seem</u> to be of at least equal significance, are more difficult to assess than wastage. Ratings by supervisors and peers are an example. Information of this kind is relatively accessible, since it can be gathered directly by interview or other techniques, collected from records, or inferred from positions

of responsibility, appointive or elective, held by a teacher. Such information, however, is subject to a number of limitations. Procedures and criteria for evaluating teachers vary from district to district, and frequently the evidence on which ratings are based is very meager or second hand. (NEA, 1969) The same problems apply to reports of changes in teachers and teacher growth. (Turner, 1965) The personality of the principal also seems to have a substantial effect on the ratings of a teacher's ability and social competence. (Start, 1948, Wiseman and Start, 1965, Wandt, 1954, Fink, 1953) In addition, school district and college supervisors do not agree in their ratings of teachers. (Start, 1967, 1) Perhaps the attempt to divide teachers into types based on profiles of attributes they have in the principal's judgment may be more valid (Johnson, M., 1965), but the evidence is not strong.

Ratings by pupils seem to be a much more promising source of information. (Remmers, H.H., 1963) Unfortunately, however, there is evidence that they do not agree with ratings by supervisors (Stern,1963), and this could be a problem from a practical point of view, making it awkward to collect the information in the schools and to explain the results. Again, there are differences in the rating of teachers by different groups. For example, older students may put more emphasis on

scholarship. (Evans, 1959)

In view of these limitations and difficulties, it does not seem possible at present to establish workable criteria based on ratings by supervisors, peers, and pupils. Other career line data are still less promising because of a lack of research studies or the inconclusiveness of the evidence so far obtained. The relationship of a teacher's participation in research and program development to his teaching appears to be unstudied. Such variables as experience, competence in the subject field, training in the teaching of that subject, further education, and so on, have been found to be related to teaching competence in some research studies but unrelated in others. Thus Blosser and Howe (1969), reviewing twenty studies, find personal adjustment and academic preparation to be related to success in teaching high school science. However, Metzner (1968) finds on reviewing seventeen research studies and reviews of research that there is no evidence of a relationship between the length of a teacher's training and his knowledge of his subject, and supervisors' ratings or pupil achievement, however measured. But administrators apparently believe that teachers should be more specifically trained for particular skills or levels of teaching - especially in stimulating thinking. (Smith M.C., 1966) And there is some evidence that those who have completed teacher preparation programs are rated as better teachers than those who have not, even when

both groups are experienced teachers (Bledsoe, 1967; Beery, 1960).

These contradictory findings may well reflect differences in the circumstances under which these factors are or are not powerful. One can easily conjecture that a thorough and sophisticated approach to a subject field, the result of superior training in the subject to be taught, could be a great asset in some classrooms and a handicap in others. The teacher's knowledge and training may be related to success in teaching the brightest students and those taking highly technical subjects in high school. (Metzner, 1968) It may be that the incidence of violence, absenteeism, students going on to college, number and type of electives offered, and so on, are variables which account for the differences in research findings. Unfortunately, researchers rarely report such details of the milieu of the schools where their projects were carried on.

Further research and improvements in the reporting of results may in time enable us to understand what the factors are that operate, but at present career data, apart from wastage, do not appear to provide a promising base for accreditation criteria.

Client satisfaction

Client satisfaction has seldom been used as a measure of the effectiveness of teacher preparation programs and is not touched on in the Recommended Standards. For several reasons, however,

it may be more suitable for accreditation procedures than most kinds of career line data. The opinions of students and graduates are a good source of subjective evidence about the effects of a program and of its various components. A program whose students hold it in high esteem is probably in a better position to affect their teaching than a program held in low esteem. Such a program is also more likely to be able to obtain information and advice from its graduates when it seeks to evaluate and improve its offerings (Lueck, 1965).

It should be noted that the arguments for using testimony of students and graduates do not necessarily hold for other groups--administrators, parents, and educational critics--whose opinions cannot be considered direct evidence of program effectiveness. Such indirect testimony may, however, have implications for program planners: for example, the evidence that most parents associate unpleasant discipline experiences with women teachers and more positive experiences with men teachers. (Lowery, 1969) For strong practical reasons, too, programs cannot afford to ignore client opinion, since it affects such matters as funding, recruitment of students, and placement of graduates.

Unfortunately, a number of problems make it difficult to conduct valid studies of client satisfaction. Students who

are still in a program or who have just begun to teach are not yet in a good position to evaluate its usefulness for teaching. After some years experience teachers do not recall the details of their training and their testimony is harder to collect. Furthermore, students' reactions to a program can vary greatly from year to year, though when these changes are in response to program changes they may be important evidence. (Herbert and Williams, 1969) The attitudes of any single group of students also seem to change during the period of their training (Fishburn, 1966), probably towards accepting the views of the teacher preparation staff and especially those of the supervising teacher. (Bloser and Howe, 1969) However, the direction of change seems to reverse itself when students graduate and begin professional teaching (Butcher, 1965, Steele, 1958), making it difficult to know when to measure client satisfaction unless these changes prove to be predictable. Graduates may however also be affected by the climate of opinion prevailing at the time of a study, as is suggested by the changes in opinions expressed about methods courses. (Albrecht, 1960, California Teachers Association, 1966) Follow-up studies have had great variation in success in getting responses, varying from 40% to over 90%, with most around the middle of this range. There is evidence to suggest that the "lost" part of the population differs from the respondents, except when the rate of response

is very high. (Start, 1967, Report I; Johnson, 1968, p. 84)

Most follow-up studies are conducted by teacher preparation staff members, who have neither the time nor the skill nor the motive to conduct rigorous analyses of the research design, procedures, and results. The problem of bias during the collection and especially during the analysis of data is high when the researcher also teaches in or administers the program under study. Of course, such research can be conducted by organizations other than the teacher preparation programs themselves, as was done by the National Union of Teachers in England. (N.U.T., 1969)

The most satisfactory base for accreditation would be a profile of the graduates' teaching derived from a set of measures of their teaching performance in a variety of appropriate situations. There are a number of practical and theoretical reasons why it is essential to include some assessment of the teaching of the graduates of a program in accreditation procedures. Changes in teacher behavior are obviously the central goal of teacher preparation. Any program that has no detectable impact on its graduates could hardly be considered effective. At the same time, information about how graduates teach is most valuable for the design and evaluation of a program by its staff, and if carried out carefully and periodically, would provide a baseline for measuring the impact of subsequent changes in the format, resources, or other variables

of the program.

The development of criteria to assess the teaching of behavior of graduates can also be justified on theoretical grounds. The analysis of teaching and of educative relationships is one of the most promising fields of educational research. The number of instruments and techniques is growing rapidly, and our knowledge is increasing both quantitatively and qualitatively. Two very useful anthologies, Mirrors of Behavior, Parts One and Two (Simon and Boyer, 1968 and 1970), give information on approximately eighty direct observation techniques, most of them developed quite recently. Without further research, unfortunately, these promising new instruments and techniques for describing, predicting, and evaluating teaching cannot be used for accreditation purposes, since they are still in the development stage. With the goal of accreditation criteria clearly in view, however, research efforts might become better coordinated and more effective.

Each type of evaluation technique has advantages and disadvantages for developing the kind of profile needed. For ease of administration and interpretation the ideal would be a test or battery of tests, with descriptive, value-free norms standardized for different populations. Work is now in progress to develop tests of this kind (McGuire and Babbott, 1967; Frederiksen, 1965) Unfortunately no test with

descriptive or predictive power for teaching has yet been developed.

Another possibility is the use of an easily staged simulated teaching environment which would make it possible to immerse the new teacher in a variety of teaching situations. Work is in progress on such situation tests, using sound film to simulate teaching sequences with facility to change the events simulated (Schalock, 1964, Kersh, 1963) and on micro-teaching and mini-lessons which provide scale models or analogues of classroom teaching (McDonald, 1967, Johnson, 1964). In this work, however, the training effect is often given more emphasis than evaluation.

While each of these procedures has an iconic relationship to actual teaching, they are isomorphic only to a limited extent. Even when they are fully developed it seems likely that a number of these situation tests would have to be combined to form a battery before one could expect much descriptive or predictive accuracy.

An alternative procedure would be to observe graduates in actual teaching situations in classrooms, laboratories, and on field trips. Until recently such observations were inevitably unsystematic, and could therefore not be used to provide precise or objective descriptions of teaching. However, as observation techniques, for example those collected in Mirrors of Behavior, become more fully developed, this drawback should cease to be

a problem. Particularly if the sample of teaching behavior is recorded, the material can be collected from a selected sample according to a previously arranged schedule and can be re-examined when evaluations disagree. Analysis of the material by means of a framework developed for the purpose can be done as needed and one of a set of suitable standards can be applied to categorize the graduates and the program.

The practical problems in the way of direct observation are much less difficult than might be anticipated. (Herbert, 1970) The theoretical problems are more serious, but these also can be resolved. Techniques of observation and analysis of teaching will have to be standardized. It would not be possible to standardize students or classroom situations, but a rough categorization of teaching situations would probably be adequate. The diversity of possible ways of teaching could make it difficult to establish a profile, but research evidence suggests that teachers actually employ a fairly limited repertoire of teaching styles. (Bellack, 1964, Foshay, 1964)

This is not the place to review the now extensive literature on the observation of teaching and learning. I have little doubt that the new edition of the Handbook o Research on Teaching will show the substantial progress which has been made since the first edition. (Gage, 1963) I believe, however, that we are now

developing techniques which, with further research, will become sufficient for describing the teaching of graduates of teacher preparation programs for purposes of accreditation. It seems to me, moreover, that teaching behavior is the ultimate criterion against wnich all other measures of the effectiveness of programs must in time be validated. The importance of the goal, I believe, should outweigh any other consideration in determining the direction of our research efforts.

Research Base

I have suggested that in establishing a research base for the accreditation of teacher preparation programs we should develop the present criteria into a set of multiple standards to fit diverse programs and several levels of quality. I have also suggested that additional standards for measuring the effects of preparation programs should be formulated. Standards serving this purpose could be based on such criteria as career lines (especially retention in teaching), client satisfaction, and, above all, the teaching behavior of students during the program and after graduation.

In preparing the present paper I gathered hundreds of papers directly or indirectly relevant to this topic. Many were discussions of criteria for accreditation or proposals for new programs of teacher preparation, often very thoughtful and ably presented; but strangely, even in the best of these discussions

there were few if any references to research. At best the authors referred to or had themselves conducted surveys of opinions and attitudes. (Stiles, 1968, AACTE, 1967) Many authors deplored the lack of research studies and spoke of teacher preparation as an "unstudied problem." Yet despite the paucity of references and the frequent call for more research, there is in fact a very large number of studies that can be drawn upon to inform discussions of accreditation and teacher preparation. Research on the description of teaching and on preparation programs dates back at least to the Commonwealth Teacher Training Study (Charters and Warples, 1929) and since that time has increased greatly, especially in the last decade. (ERIC, 1969, Lindsey, 1969, Eidell 1968, AACTE, 1968, Heidelbach & Lindsey, 1968, Canadian Teacher Federation, 1969) Can we then draw upon these studies to form a research base for the new criteria? We can, but there are some major problems.

Inadequacy of Reporting

Most of the research studies are reported in journals. A substantial number of other studies remains unpublished (though where these are doctoral dissertations they can be traced), and another large group, especially reports of follow-up studies, were never completed. Even in the published reports of research, however, much information about the procedures and the milieu of the study is usually omitted. For example, the reports rarely

state the type of school or schools in which the research was conducted or describe the educational program of the schools, the socio-economic level of the pupils, the ages and academic and professional preparation of the teachers, the teaching styles they employed, and so on, except when these were the variables directly under study. Yet these variables clearly often affect the results. A researcher who wanted to make use of a study for almost any purpose would need at the very least to go back to the primary research report, and perhaps even to contact the investigators before he couid interpret the results.

The lack of replication is also a very serious problem. Replications of research studies are as rare as reports of peace in the newspapers. Strangely, researchers will often produce a single instance as though it were generalizable. Their next piece of research is usually quite different, and no one verifies any results, so that there is no evidence that another experimenter or the same experimenter at another time or place would have obtained the same results. The erratic distribution of research topics is still another problem. Dussault (1969) reports that he found sixteen studies of the effect of supervision on the attitudes of student teachers, and only one study (Brown, 1962) of the effect of supervision on their teaching behavior. This situation is quite widespread, with the result that no research at all has been done

in some key areas.

Heavy reliance on theories external to pedagogy is another problem. Ideally, research studies can be placed into a network of theory which relates them to one another. The lack of a common theoretical framework means not only that research results often are not directly comparable, but also that it is very difficult to conduct a rigorous review of research, since it is necessary to analyze the terms used and check them against the theory in which they are imbedded in order to interpret the methodology of a study and its results.

Lack of Screening

Perhaps this difficulty is the main reason why these studies are rarely examined and tested rigorously, along the lines of some recent correspondence. (Rosenthal et al., 1968, Thorndike, 1968, Rosenthal, 1969, Thorndike, 1969) In the absence of such uniform procedures as are found in the natural sciences, the likelihood of error is very high, and the resulting errors may hide significant results or produce a deceptive significance. If enough studies are conducted the probability of some statistically significant results occuring at random is high. As studies with statistically significant results are those most likely to be reported, distortion of information is very likely.

When these problems are considered, it is clear that some

rigorous analysis of research results is needed before the findings can be incorporated in a research base.

Proposals

One step toward the preparation of a research base which could serve the design and maintenance of teacher preparation programs as well as accreditation procedures would be the establishing of an evaluation team for screening research relevant to teacher preparation. Such a team should include experts in research design and in teacher preparation as well as generalists who can take an overall view.

It seems unrealistic to expect each teacher preparation program to design, initiate, conduct, and analyze its own research studies, as Section 5 of the present Recommended Standards seems to require. Such studies would be seriously handicapped by a lack of the trained staff, the resources, and the mental set and orientation necessary for independent research. The duplication of effort would in any case be highly wasteful.

It is even doubtful whether any single organization--a teacher preparation program, or this Special Interest Group on the Teacher Preparation Curriculum, or the NEA, or Division B of the AERA--could marshall adequate resources to establish a research base for accreditation. But by working in cooperation, these and other groups could develop a number of varied but rigorous and

interconnected research designs, with procedures for processing the data. To such a collaborative effort, the teacher preparation programs could contribute the special knowledge and resources which they possess. The participation of researchers not connected with the programs would ensure careful design, rigorous procedures, and a minimum of bias. Replication could be ensured by making the same standardized research designs available to many programs. Variety and scope could be ensured by providing a number of different designs. Teacher preparation programs would benefit from accurate feedback about their graduates and from the economy with which they would obtain such information.

In this way we could build a solid research base for accreditation on replicated, carefully designed studies. Given the possibility of cooperative research, there seems to be no good reason why 1100 teacher preparation programs in the United States, and hundreds more in the English speaking world, should have to design their own research and follow-up programs. Every teacher preparation institution would, with relief, participate in a project which promised to help rather than to police its program.

Bibliography

Albrecht, G. A survey of teacher opinion in California. Phi Delta Kappan, 1969, 42, 103-108.

American Association of Colleges for Teacher Education. Research and study in teacher education. AACTE Bulletin, 1968, 20 (9).

American Association of Colleges for Teacher Education. Standards for the accreditation of teacher education. (2nd draft) Washington, D.C.: AACTE, 1968.

American Association of Colleges for Teacher Education. Recommended standards for teacher education. Washington, D.C.: AACTE, 1969.

Beery, J. R. Professional preparation and effectiveness of beginning teachers. Coral Gables, Florida: Graphic Arts Press, University of Miami 1960.

Bellack, A. A., Kliebard, H. M., Hyman, R. T., & Smith, F. L., Jr. The language of the classroom. New York: Teachers College Press, Columbia University, 1966.

Bledsoe, F., et al. Comparison between selected characteristics and performance of provisionally and professionally certified beginning teachers in Georgia. Final report. Athens, Georgia: Georgia University, 1969.

Blosser, P. E., & Howe, R. W. An analysis of research related to the education of secondary school science teachers. The Science Teacher, 1969, 36, 87-95.

Brase, P. C., Jr. A study of curriculum changes in selected two year colleges initiated by the evaluation process used by the Middle States Association. New York: Teachers College, Columbia University, 1964.

Brown, B. B. An investigation of observer-judge ratings of teacher competence. Final report. Gainesville, Florida: Florida University, 1969.

Bruce, M. H. Teacher education in science. Review of Educational Research, 1969, 39, 415-427.

Butcher, H. J. The attitudes of student teachers to education: A comparison with the attitudes of experienced teachers and a study of changes during training courses. British Journal of Social and Clinical Psychology, 1965, 4(Feb.), 17-24.

California Teachers' Association. California teachers' reactions to certification and pre-service courses. Supplementary Research Report No. 58. Burlingame: Department of Research, CTA, 1966.

Canadian Teachers' Federation. Pre-service teacher education in Canada. Bibliographies in Education No. 1. Ottawa: The Federation, 1969.

Charters, W. W., Waples, D., & Capen, S. P. The commonwealth teacher-training study. Chicago: University of Chicago Press, 1929.

Charters, W. W., & Waples, D. The commonwealth teacher-training study. Chicago: University of Chicago Press, 1929.

Dussault, G. A theory of supervision in teacher education. New York: Columbia University, Teachers College Press, in press.

Eidell, T., & Klebe, J. Annotated bibliography on the evaluation of educational programs. Eugene, Oregon: Oregon University, 1968.

Evans, K. M. Research on teaching ability. Educational Research, 1959, 1, 22-36.

Evers, N. H. By what yardstick shall be judged? Evaluative criteria for accrediting teacher education: A source book on selected issues. Washington, D.C.: American Association of Colleges for Teacher Education, 1967, Pp. 59-67.

Fishburn, C. E. Learning the role of the teacher. Journal of Teacher Education, 1966, 17, 329-331

Fink, M. Ethnocentrism as it relates to teaching success. California Journal of Educational Research, 1953, 4, 111-114.

Foshay, A.W. Curriculum and instruction. In Dundee team teaching project, Greenwich, Connecticut: A research report. New York: Institute of Field Studies, Teachers College, Columbia University 1965, Pp. 71-176.

Frederiksen, N. Proficiency tests for training education. In R. Glaser (Ed.), Training research and education. New York: Science Editions, 1965, 323-346.

Heidelbach, R. & Lindsey, M. Annotated bibliography on laboratory experiences and related activities in the professional education of teachers, July 1966 - June 1967. Washington, D.C.: Association for Student Teaching, 1968.

Herbert, J. Direct observation as a research technique. Psychology in the Schools, 1970, in press.

Herbert, J., & Williams, D. P. Psychology in the teacher preparation program. In J. Herbert & D. P. Ausubel (Eds.), Psychology in teacher preparation. Monograph Series No. 5. Toronto: Ontario Institute for Studies in Education, 1969.

Johnson, J. A. A national survey of student teaching programs. Final report, 1968, Northern Illinois University, Bureau No. BR-6-8182, Bureau of Research, Office of Education, D.H.E.W., Washington, D.C.

Johnson, M., Jr. Profiles of beginning junior high school teachers. Journal of Teacher Education, 1965, 16, 303-306.

Johnson, W. D. Microteaching: A medium in which to study teaching. The high school journal, 1967, 51(2), 96-92.

Kersh, B. Classroom simulation: A new dimension in teacher education. Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1963.

Lindsey, M., et al. Annotated bibliography on the professional education of teachers. Washington, D.C.: Association For Student Teaching, 1969.

Lowery, L. F. A study of the attitudes of parents toward teachers. Journal of Educational Research, 1969, 62, 227-230.

Lueck, W. R. Professional insecurity of prospective teachers. Teachers College Journal, 1965, 38, 119-125.

MacDonald, J. B. Researching curriculum output: The use of a general systems theory to identify appropriate curriculum outputs and research hypotheses. Paper presented at the annual meeting of the American Educational Research Association, 1965.

McDonald, F. J., & Allen, D. W. Training effects of feedback in modelling procedures on teaching performance. Stanford, California: Department of Education, Standford University, 1967.

McGuire, C. H. A process approach to the construction and analysis of medical examinations. Journal of Medical Education, 1963, 38, 556-563.

McGuire, C. H., & Babbott, D. Simulation technique in the measurement of problem-solving skills. Journal of Educational Measurement, 1967, 4(1), 1-10.

Maucker, J. W. Imperatives for excellence in teacher education. Foundations for excellence: Fifteenth Yearbook, American Association of Colleges for Teacher Education, Washington, D.C.: AACTE, 1962.

Metzner, S. The teacher preparation myth: A phoenix too frequent. Phi Delta Kappan, 1968, 50, 105-107.

National Education Association, Research Division. Teacher mobility and turnover, 1965-66 to 1966-67. Washington, D.C.: National Education Association, 1969.

National Education Association. Teacher mobility and loss. National Education Association Research Bulletin, 1968, 46, 118-126.

National Union of Teachers. The future of teacher education. London: National Union of Teachers, 1969.

Orlich, D. C. Idaho teacher turnover: 1965 - A selected analysis of the problem. Journal of Teacher Education, 1967, 17, .447-453

Remners, H. H. Rating methods in research on teaching. In N.L. Gage, Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 329-378.

Rosenthal, R. Empirical versus decreed validation of clocks and tests. American Educational Research Journal, 1969, 6(4), 689-691.

Rosenthal, R., & Jacobson, L. Pygmalion in the classroom. New York: Holt, Rinehart & Winston, 1968.

Ryans, D. G. Patterns of teacher classroom behavior and their assessment. Characteristics of teachers: Their description, comparison, and appraisal. Washington, D.C.: American Council on Education, 1960.

Schalock, H. D., Beaird, J. H., & Simmons, H. Motion pictures as test stimuli: An application of new media to the prediction of complex behavior: Monmouth, Oregon: Teaching Research Division, Oregon State System of Higher Education, 1964.

Simon, A., & Boyer, E. Gil. (Eds.) Mirrors for behavior: An anthology of classroom observation instruments. Vols. 1-6. Philadelphia, Penn.: Research for Better Schools, Inc., 1968.

Simon, A., & Boyer, E. Gil. (Eds.) Mirrors for behavior: An anthology of classroom observation instruments continued. Vols. 7-15. Philadelphia, Penn.: Research for Better Schools, Inc., 1970.

Smith, M. C. The case for teachers who are specifically prepared to teach in junior high schools. Journal of Teacher Education, 1966, 17, 438-443.

Start, K. B. A follow-up of the 1961 group of teachers. Selection, training and professional progress of teachers, Report No. 1. Manchester University, School of Education, 1967.

Start, K. B. Teachers from different colleges of education. Selection, training and professional progress of teachers, Report No. 2. Manchester University, School of Education, 1967.

Steele, P. M. Changes in attitude amongst training college students towards education in junior schools. Unpublished masters thesis, University of Manchester Library, 1958.

Stern, G. Measuring non-cognitive variables in research on teaching. In N. L. Gage, Handbook of research on teaching. Chicago: Rand McNally, 1963, Pp. 398-447.

Stiles, L. J., et al. Teacher certification and preparation in Massachusetts. Boston: Massachusetts Advisory Council on Education, 1968.

Thorndike, R. L. Review of R. Rosenthal and L. Jacobson, Pygmalion in the classroom. American Educational Research Journal, 1968, 5, 708-711.

Thorndike, R. L. But you have to know how to tell the time. American Educatonal Research Journal, 1969, 6(4), p. 692.

Turner, R. L. Characteristics of beginning teachers: Their differential linkage with school system types. School Review, 1965, 73, 48-58.

Wandt, E. A comparison of the attitudes of contrasting groups of teachers. Educational and Psychological Measurement, 1954, 14(2), 418-422.

Wiseman, S., & Start, K. B. A follow-up of teachers five years after completing their training. British Journal of Educational Psychology, 1965, 34, 342-361.

Woodring, P. New directions in teacher education: An interim report of the fund for the advancement of education in the areas of teacher education and recruitment. New York: Fund for the advancement of education, 1957.